

STEWARDSHIP OF VERY LARGE DIGITAL DATA ARCHIVES

Patric Savage
Shell Development Company

Call an archive a permanent store,

Some of the largest *digital data archives* are operated by oil exploration organizations. The vast bulk of these archives is seismic data. It is kept forever because of its extremely high cost of acquisition, and because it often cannot be re-acquired (due to cultural buildup, political barriers, or difficult logistical/administrative factors). Western Geophysical operates a seismic data archive in Houston consisting of more than 725,000 reels/cartridges, typical of the industry. Oil companies fondly refer to their seismic data troves as "family jewels."

There are relatively few "very large" digital data archives in existence. Most business records are (gladly) expired within five or ten years depending on statutes of limitations. And many kinds of business records that do have long lives are embedded in data bases that are continually updated and re-issued cyclically. Also a great deal of "permanent" business records are actually archived as microfilm, fiche, or optical disk images - their digital version being an operational convenience rather than an archive.

So there is not really much widely known about operating *digital data archives*, let alone *very large* ones. Even the oil companies have been in a sense overwhelmed by this somewhat unplanned for hugeness.

This paper addresses the problems foreseen by the author in stewarding the very large digital data archives that will accumulate during the mission of the EOS. It focuses on the function of "shepherding" archived digital data into an endless future.

Stewardship entails a great deal more than storing and protecting the archive. It also includes all aspects of providing meaningful service to the community of users (scientists) who will want to access the data. The complete steward will:

1. Provide against loss due to physical phenomena.
2. Assure that data is not "lost" due to storage technology obsolescence.
3. Maintain data in a current formatting methodology. Also, it may be a requirement to be able to reconstitute data to original as-received format.
4. Secure against loss or pollution of data due to accidental, misguided, or willful software intrusion.

5. Prevent unauthorized electronic access to the data, including unauthorized placement of data into the archive.
6. Index the data in a *metadatabase* so that all anticipatable queries can be served without searching through the data itself.
7. Provide responsive access to the *metadatabase*.
8. Provide appropriately responsive access to the data.
9. Incorporate additions and changes to the archive (and to the *metadatabase*) in a timely way.
10. Deliver only copies of data to clients - retain physical custody of the "official" data.

Items 4 through 10 are not discussed in this paper. However, the author will answer questions about them at the conference or by email or telephone.

Providing Against Loss Due to Physical Phenomena

Broadly classifying these we have:

1. Site destruction
2. Theft/robbery
3. Sabotage
4. Media unit suffers severe damage
5. Systemic media degradation

The first three can be guarded against, but not absolutely. The fourth is a rare inevitable eventuality (e.g., a mechanically faulty drive "eats" a tape.)

Systemic media degradation is best managed by using only media that are known to have archival properties, by conservatively rewriting media that, when accessed, are found to have an error, by regularly running PM according to vendors' recommended practice (e.g., winding and re-tensioning tape), and copying the entire archive to new-generation media. The last must be planned for, budgeted for, and be resigned to - it is an imperative. Generally speaking, one media generation can be leapfrogged by the copy procedure: for example, when Shell adopted 3480 technology, all of the 1100bpi tapes were copied; when 3490 technology is adopted, all of the 6250bpi tapes will be copied. However, copying can be mandated earlier if media are observed to be systemically degrading faster than anticipated.

Media failure occurs when an uncorrectable bit error is detected. This always causes/implies loss of an entire error correction block - ordinarily a minimum of one kilobyte of archived data. Archivists should be aware that the media vendors' touted "hard error rate" always has a 10,000-fold impact. A badly degraded media unit might have relatively many unreadable error correction blocks; hence even a redundancy array of media units might then (by a little bad luck) have an unrecoverable error correction block.

A practical cost-effective solution to the problem of protecting against physical loss can be tailored around the following concept (which came to me while ruminating about extending the now-familiar RAID idea to striping tape). This is merely the seed of an idea, to which a good deal of systems thought will have to be given.

Some number (in this example, 10) of archiving sites are chosen to participate/cooperate in a redundancy scheme that provides mutual protection against all modes of physical loss.

The sites should be geographically distributed in order to eliminate concern that a calamity (e.g., earthquake/meteor strike) would wipe out multiple archives. Of course, all sites individually should have reasonably good physical security.

All sites must be accessible via state-of-the-art WAN technology.

Each site houses, primarily, its own archive of data. (A variation would have a single archive partitioned and distributed among its own multiple sites.) Clients of an archive would communicate only with the primary site.

Each site also houses either p-parity or q-parity data generated from (in this example) 9 other sites. (Optionally two sites could be dedicated, one for p- and the other for q-parity.)

In the eventuality of a loss at a site (of an error correction block, or a media unit, or the site itself) any 8 of the 9 other sites reconstruct the lost data. This would not be instantaneous, as with RAID, because an extraordinary procedure would have to be executed; but the insurance would be very certain. Clearly, each site should be practicing high quality archiving methodology, so that losses would occur with extreme rarity (say, no oftener than one per month).

The merits of this scheme are first, that the storage overhead for backup can be small (25% for this example); second, that the degree of protection can be high (with both p- and q-parity) or lower (with p-parity only); third, independent archives do not have to create their own backup systems, but can band together in a consortium for mutual protection.

Assuring That Data is Not "Lost" Due to Storage Technology Obsolescence

The 1960 census was archived on the best storage medium known at the time: UNIVAC metal tape. There was a rude awakening some years later when it was discovered that only two drives existed in the world - one in Japan, and the other, dismantled, in the Smithsonian.

We know now that drive technology lifetime, even assuming heroic geriatric care, is scarcely ten years. Vendors drop maintenance after low-level parts' technologies disappear. For a while thereafter, drives can be cannibalized for parts; but ultimately maintenance becomes impossible.

The optical disk vendors, for example, tell of the fine archival qualities of their media. But their technology is evolving quite rapidly - vendors come and go, and recording formats with them. Here we have a single medium that undoubtedly lasts a long time, but the drive and recording technology has a half-life of less than five years. Considering the relatively high cost of optical media, copying an archive every five years seems out of reason.

Archiving demands that digital data on old storage technology be copied to new storage technology periodically. The frequency depends on the media, on how widely the drives were accepted, and on whether the old technology satisfies current access requirements. Keeping too many generations of storage technology in use can cause serious operational problems, even if they are all in good working condition. For example, 556bpi tape would be much too slow for regular use today, so, even though drives are still available, that technology is obsolete.

Maintaining Data in a Current Formatting Methodology

The winds of computation methodology are ever varying. Yesterday there was no C language. Today C-readable records might be a good bet. Tomorrow the fad may be object files. What will come next? The curse of required media copying is really a blessing because it enables us to continually modernize our data language. Cuneiform tablets were certainly archival, but they contain antiquated, almost unreadable language.

Standards for "self-defining" data formats are evolving rapidly and are already very useful. The time has come to abandon schema-less data formats (where programs know implicitly where every field is in a record, and what each field means).

Even fixed (schema'd) formats are passé for scientific data because of the continual change in interest and emphasis in almost every scientific specialty.

Archivists can extract a side benefit when copying to a new media generation. Indeed, the planning for the copy should include deciding which new formatting standard is to be adopted. Migrating from old to new formats is only slightly less important for archiving as migrating from old to new media technology. What's more, it's almost free.

High-Performance Storage Systems

**Robert Coyne
IBM Federal Systems Research Division**

**Permission for this talk could not be obtained in time for the Text to be included in
these proceedings.**

PRECEDING PAGE BLANK NOT FILMED

